

Analisis Survival dengan Pendekatan Multivariate Adaptive Regression Splines pada Kasus Penyakit Ginjal Kronis

Hajar Aziyah Mutiah¹, Joko Purwadi²

¹*Program Studi Matematika, Fakultas Sains dan Teknologi Terapan, Universitas Ahmad Dahlan*

Email: hajar1500015032@webmail.uad.ac.id

²*Program Studi Matematika, Fakultas Sains dan Teknologi Terapan, Universitas Ahmad Dahlan*

Email: Joko@math.uad.ac.id

ABSTRAK

Analisis Survival merupakan metode statistik yang berkaitan dengan waktu, mulai dari *time origin* atau *start point* sampai dengan terjadinya suatu kejadian khusus atau *end point*. Data *survival* yang melibatkan banyak variabel prediktor dapat dianalisis dengan regresi. Regresi yang digunakan untuk menganalisis data *survival* yaitu regresi parametrik, regresi nonparametrik, dan regresi semiparametrik. Penelitian ini menggunakan salah satu regresi nonparametrik yaitu *Multivariate Adaptive Regression Splines*. *Multivariate Adaptive Regression Splines* (MARS) merupakan pendekatan untuk regresi multivariat nonparametrik yang menghasilkan pemodelan regresi yang fleksibel. Metode ini memiliki kelebihan fleksibel pada data berdimensi tinggi. Analisis *survival* dengan pendekatan *Multivariate Adaptive Regression Splines* (MARS) untuk mencari faktor faktor yang mempengaruhi laju kegagalan pasien diterapkan pada kasus Penyakit Ginjal Kronis berdasarkan data rekam medis di Rumah Sakit PKU Muhammadiyah Bantul. Dari penelitian ini diperoleh permodelan Cox dengan pendekatan MARS yang menunjukkan bahwa variabel yang berpengaruh terhadap laju ketahanan hidup pasien Penyakit Ginjal Kronis yaitu Usia(X_1), Tekanan Darah Sistolik(X_3) Kadar Ureum(X_5) dan Kadar Hemoglobin(X_7).

Kata kunci : *Analisis Survival, Martingale Residual, Multivariate Adaptive Regression Splines, Penyakit Ginjal Kronis.*

PENDAHULUAN

Analisis Survival merupakan salah satu analisis yang menggambarkan suatu proses yang berhubungan dengan waktu, diawali dengan *time origin* atau *start time* sampai dengan terjadinya suatu kejadian khusus atau *end point*. Analisis *survival* memiliki dua fungsi yang dapat digunakan, yaitu fungsi *survival* dan fungsi *hazard*. Dalam analisis *survival* diasumsikan hanya ada satu kejadian pada obyek yang diteliti. Untuk melihat hubungan variabel respon dan variabel prediktor pada data *survival* maka dapat digunakan regresi.

Multivariate Adaptive Regression Splines (MARS) merupakan salah satu metode nonparametrik yang fleksibel untuk pemodelan regresi dengan data berdimensi

tinggi. Permodelan MARS menunjukkan interaksi antar variabel prediktor dan mampu mendeteksi interaksi tersebut. Pada analisis survival dengan pendekatan MARS residual *martingale* dari model regresi Cox Proportional Hazard digunakan sebagai variabel respon dari pemodelan MARS.

Penelitian sebelumnya tentang analisis *survival* dengan pendekatan MARS telah dilakukan oleh Kriner (2007) pada kasus ketahanan hidup pasien jantung di Jerman, dimana metode MARS menunjukkan hasil yang lebih baik daripada regresi Cox. Metode yang sama juga telah diterapkan oleh Nisa dan Budiantara pada kasus Demam Berdarah Dengue di Kabupaten Gresik dimana variabel yang mempengaruhi laju kesembuhan pasien yaitu jumlah trombosit, kadar hematokrit, usia dan perbesaran hati.

Penyakit ginjal kronis adalah suatu proses patofisiologis dengan etiologi yang beragam, mengakibatkan penurunan fungsi ginjal yang progresif, dan pada umumnya berakhir dengan gagal ginjal. Pada kondisi Penyakit Ginjal Kronis ginjal tidak dapat menyaring kotoran, tidak mampu mengontrol jumlah air dalam tubuh serta kadar garam dan kalsium dalam darah. penyebab penyakit ginjal kronis antara lain penyakit infeksi, penyakit peradangan, penyakit vaskuler hipertensif, gangguan jaringan ikat, gangguan kongenital dan hereditas, penyakit metabolik, nefropati toksik, nefropati obstruktif.

Penyakit Ginjal Kronis merupakan masalah global yang jumlah penderitanya semakin meningkat. Seperti yang dicatat oleh *Indonesia Renal Registry* (IRR) di tahun 2016 pasien PGK mencapai 21.050. Jumlah tersebut meningkat menjadi 25.446 pada tahun 2017. Sementara itu data Ikatan Dokter Anak Indonesia (IDAI) menyebutkan pada tahun 2017 terdapat 212 anak dari 19 rumah sakit di Indonesia yang mengalami gangguan ginjal dan menjalani cuci darah, sedangkan angka kematian pada anak yang mengalami gagal ginjal yaitu 23,6 persen.

Penelitian ini bertujuan mengetahui faktor-faktor yang berpengaruh terhadap laju kegagalan hidup pasien Penyakit Ginjal Kronis. Batasan masalah pada penelitian ini yaitu data yang digunakan berupa data rekam medis pasien yang menjalani pengobatan

di Rumah Sakit PKU Muhammadiyah Bantul sejak bulan Juli 2018 sampai dengan bulan Juli 2019. Pasien yang diteliti adalah pasien yang menjalani pengobatan hingga dinyatakan meninggal. Manfaat dari penelitian ini bagi instansi dan masyarakat yaitu memberikan informasi mengenai faktor-faktor yang paling mempengaruhi probabilitas ketahanan hidup pasien PGK. Kemudian manfaat bagi mahasiswa yaitu agar memahami penerapan ilmu statistika di bidang kesehatan khususnya penggunaan analisis *survival* dengan pendekatan MARS.

LANDASAN TEORI

Analisis *Survival*

Analisis *survival* adalah suatu metode statistik yang berkaitan dengan waktu, dimulai dari *time origin* atau *start poin* sampai pada suatu kejadian khusus (*failure event/end point*). Dalam menentukan waktu *survival* T , terdapat tiga elemen yang perlu diperhatikan yaitu :

1. *Time origin* atau *starting point* (titik awal) adalah waktu dimuali nya suatu penelitian.
2. *Ending event of interest* (kejadian akhir) adalah kejadian yang menjadi inti dari penelitian.
3. *Measurement for the passage of time* (skala pengukuran waktu yang jelas). skala diukur dalam hari, minggu atau tahun.

Fungsi *Survival* dan Fungsi *Hazard*

Pada analisis *survival* terdapat dua fungsi utama, yaitu fungsi *survival* dan fungsi *hazard*. Persamaan fungsi *survival*, jika T melambangkan waktu *survival*, dan $S(t)$ merupakan probabilitas waktu *survival* lebih besar dari t , maka persamaannya adalah sebagai berikut :

$$S(t) = P(T \geq t) = 1 - F(t)$$

dimana $F(t)$ merupakan fungsi distribusi kumulatif dari T . Fungsi *survival* digunakan untuk mengetahui probabilitas waktu *survival* dari waktu mulai (*starting point*) hingga waktu ke- t . Fungsi *hazard* $h(t)$ adalah laju *failure* atau kegagalan suatu pasien untuk mengalami kejadian dalam interval waktu dari t sampai Δt dengan syarat pasien telah bertahan sampai waktu ke- t .

$$h(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t < T \leq t + \Delta t | T \geq t)}{\Delta t}$$

Hubungan antara fungsi *survival* dan fungsi *hazard* adalah sebagai berikut :

$$h(t) = \frac{f(t)}{S(t)}$$

Analisis Kaplan-Meier

Analisis Kaplan Meier adalah metode nonparametrik yang digunakan untuk menaksir fungsi *survival*. Kurva *survival* Kaplan-Meier adalah suatu kurva yang menggambarkan hubungan antara estimasi fungsi *survival* pada waktu t dengan waktu *survival*.

$$\hat{S}(t) = \begin{cases} 1 & \text{jika } t \leq t_1 \\ \prod_{j:t_j \leq t} \left(1 - \frac{d_j}{n_j}\right) & \text{jika } t_i \leq t \end{cases}$$

dimana :

d_j : banyaknya event

n_j : banyaknya individu yang berisiko

Uji Log-Rank

Uji Log-Rank adalah uji hipotesis untuk menguji secara statistik perbedaan pada kurva *survival* Kaplan-Meier antara dua kelompok data atau lebih. Hipotesis yang digunakan pada uji Log-Rank untuk dua atau lebih kelompok adalah sebagai berikut:

H_0 : Tidak ada perbedaan kurva *survival* antara grup yang berbeda

H_1 : Minimal ada satu perbedaan kurva *survival* antara grup yang berbeda.

Statistik uji pada uji log rank adalah:

$$\chi^2 = \sum_{i=1}^G \frac{(O_i - E_i)^2}{E_i}$$

dimana :

O_i : nilai observasi individu grup ke - i

E_i : nilai ekspektasi individu grup ke - i

Model Cox Proportional Hazard

Model *Cox Proportional hazard* adalah pemodelan metode semiparametrik yang digunakan untuk mengestimasi efek *covariate* pada data *survival*.

$$h(t, x) = h_0(t) \exp \left(\sum_{i=1}^p \beta_i x_i \right)$$

dimana :

$h(t, x)$: banyaknya event

$h_0(t)$: fungsi baseline hazard

β_i : parameter model regresi, dengan $i=1,2,\dots, i = 1,2,\dots, p$

x_i : variabel prediktor, dengan $i=1,2,\dots, i = 1,2,\dots, p$

Dengan menerapkan model regresi *Cox*, maka akan diketahui bentuk hubungan antar variabel bebas dan variabel tak bebas pada data *survival*.

Pemodelan dengan menggunakan *Cox Proportional Hazard* menghasilkan beberapa jenis residual, salah satunya adalah *Martingale Residual*. Dalam penelitian ini digunakan *Martingale Residual* yang berfungsi sebagai variabel respon untuk pemodelan MARS. Persamaan *Martingale Residual* adalah sebagai berikut :

$$M_i(t) : N_i(t) - H_i(t)$$

dimana :

$M_i(t)$: residual martingale ke-i pada waktu ke-t

$N_i(t)$: 1 untuk data tidak tersensor, 0 untuk data tersensor

$H_i(t)$: Fungsi hazard kumulatif

Multivariate Adaptive Regression Splines

Multivariate Adaptive Regression Splines (MARS) merupakan pendekatan untuk regresi multivariat nonparametrik yang menghasilkan pemodelan regresi yang fleksibel. MARS

merupakan bentuk perluasan dari Basis Fungsi Splines dimana jumlah Basis Fungsi merupakan parameter dari pemodelan tersebut.

Beberapa istilah yang perlu diperhatikan dalam metode dan pemodelan MARS adalah sebagai berikut,

1. Knots

Knots merupakan titik dari sebuah garis regresi untuk membentuk region dari suatu fungsi regresi.

2. Basis Fungsi (BF)

Merupakan kumpulan dari beberapa fungsi yang digunakan untuk menjelaskan hubungan antara variabel respon dan variabel prediktor.

3. Interaksi

Merupakan hubungan korelasi antar variabel dengan jumlah Maksimum Interaksi (MI) adalah 1, 2, dan 3.

Persamaan umum model MARS adalah sebagai berikut

$$\hat{f}(x) = \alpha_0 + \sum_{K_m=1} f_i(x_i) + \sum_{K_m=2} f_{ij}(x_i, x_j) + \sum_{K_m=3} f_{ijk}(x_i, x_j, x_k) + \dots$$

Persamaan diatas menunjukkan penjumlahan pertama untuk semua basis fungsi yang mengandung satu variabel prediktor. Penjumlahan kedua untuk semua basis fungsi yang mengandung dua variabel prediktor, untuk interaksi antara dua variabel dan seterusnya.

Pemodelan MARS ditentukan berdasarkan *trial and error* untuk kombinasi BF, MI, dan MO untuk mendapatkan nilai GCV yang minimum. Persamaan GCV adalah sebagai berikut :

$$GCV = \frac{\frac{1}{n} \sum_{i=1}^n [y_i - \hat{f}_{M(x_i)}]^2}{\left[1 - \frac{C(M)}{n}\right]^2}$$

Dengan :

$C(M) = \text{trace} (B(B^T B)^{-1} B^T) + 1 =$ banyaknya parameter yang diestimasi

$n =$ Banyaknya data

$B =$ Basis fungsi

$M =$ Jumlah basis fungsi

$d = 2 \leq d \leq 4$

$\frac{1}{n} \sum_{i=1}^n [y_i - \hat{f}_{M(x_i)}]^2 =$ Average Sum Square of Residual (ASR)

METODE PENELITIAN

Sumber Data

Data yang digunakan adalah data rekam medis pasien Penyakit Ginjal Kronis (PGK) yang menjalani pengobatan di Rumah Sakit PKU Muhammadiyah Bantul pada bulan Juli 2018 sampai dengan Juli 2019 sebanyak 148 data.

Variabel Penelitian

T : Waktu *survival*

d : Status Sensor (0 = Tersensor dan 1 = Meninggal)

x_1 : Usia

x_2 : Status Anemia (1= Anemia, 2 = Tanpa Anemia)

x_4 : Tekanan Darah Sistolik (1 = < 130 , 2 = 130-150 , 3= >150) mmHg

x_5 : Tekanan Darah Diastolik (1 = <80 , 2 = \geq 80) mmHg

x_6 : Kadar Ureum (1= <80, 2 = 80-120, 3 = 120-150, 4= >150) mg dL

x_7 : Kadar Kreatinin (1 = \leq 5, 2 = 5,1-10, 3=10.1-15, 4=>15) mg dL

x_8 : Kadar Hemoglobin (1 =<7, 2 = 7-10, 3=>10) mg dL

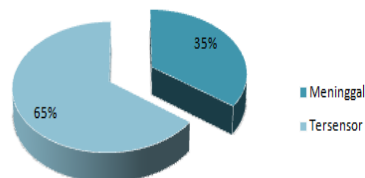
Langkah Penelitian

1. Melakukan analisis Kaplan Meier dan Uji Log-Rank.
2. Membentuk model *Cox Proportional Hazard*.
3. Membuat *scatterplot martingale residual* dengan variabel usia.
4. Menggunakan *martingale residual* sebagai variabel respon pada permodelan MARS
5. Menentukan kombinasi Basis Fungsi , Maksimum Interaksi dan Minimum Observasi
6. Mendapatkan model terbaik dari nilai GCV dan MSE
7. Memodelkan *Cox PH* dengan MARS
8. Menginterpretasi model Analisis Survival dengan pendekatan MARS yang telah didapatkan.

HASIL DAN PEMBAHASAN

Karakteristik Pasien Penyakit Ginjal kronis

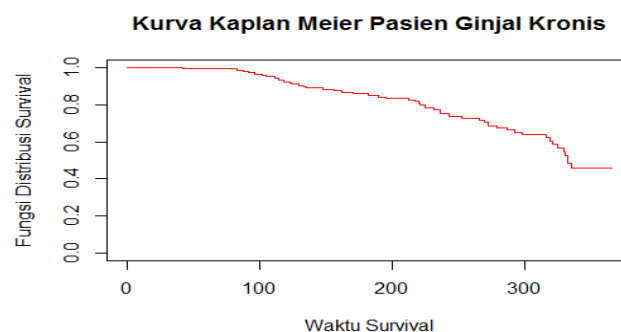
Status Pasien



Gambar diatas menunjukkan berdasarkan karakteristik waktu *survival* dapat diketahui bahwa dalam penelitian ini lebih banyak terdapat data tersensor dibandingkan data tidak tersensor yaitu sebanyak 65%. Sedangkan untuk pasien berstatus meninggal sebanyak 35%.

Analisis Kaplan-Meier

Analisis Kaplan-Meier digunakan untuk menaksir fungsi *survival*. Selain itu karakteristik waktu *survival* pasien ginjal kronis dapat ditunjukkan dengan menggunakan kurva *survival* Kaplan-Meier. Berikut ini adalah kurva *survival* Kaplan Meier secara keseluruhan untuk mengetahui gambaran karakteristik pasien secara umum.



Dari gambar diatas dapat disimpulkan bahwa angka probabilitas ketahanan hidup semakin bertambahnya hari semakin kecil.

Uji Log-Rank

Untuk menguji hipotesis apakah terdapat perbedaan antara kurva *survival* pasien ginjal kronis berdasarkan faktor-faktor yang diduga mempengaruhi ketahanan hidup

pasien ginjal kronis, maka perlu dilakukan uji Log-rank pada masing-masing variabel. Berikut hasil uji Log-rank untuk tiap variabel.

Variabel	Log-rank	P-value
Usia	359	2×10^{-16}
Status Anemia	11	9×10^{-4}
Tekanan Darah Sistolik	73,3	2×10^{-16}
Tekanan Darah Diastolik	11,1	8×10^{-4}
Kadar Ureum	110	2×10^{-16}
Kadar Kreatinin	105	2×10^{-16}
Kadar Hemoglobin	26,3	2×10^{-6}

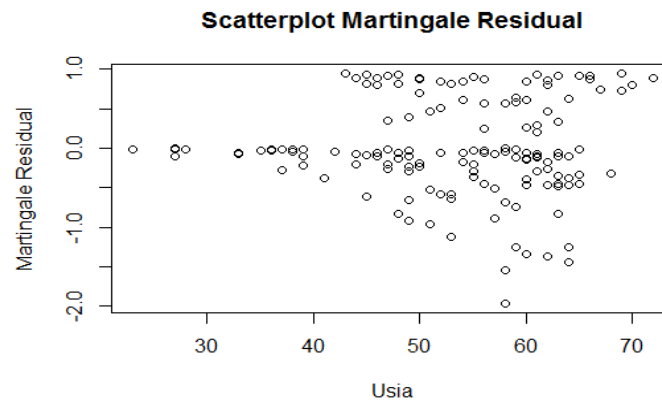
Dari tabel diatas dapat semua variabel yang diduga mempengaruhi ketahanan hidup pasien ginjal kronis memiliki nilai $p\text{-value} < 0,05$,sehingga dapat disimpulkan semua variabel memiliki perbedaan kurva *survival*.

Model Cox Proportional Hazard

Persamaan regresi *Cox Proportional Hazard* merupakan model berdistribusi semi-parametrik karena dalam persamaan Cox tidak perlu memerlukan informasi mengenai distribusi khusus yang mendasari waktu *survival* dan untuk mengestimasi parameter regresi Cox tanpa harus menentukan fungsi *hazard baseline*. Berikut model regresi *Cox Proportional hazard* dengan bantuan software R :

$$\begin{aligned} \hat{h}(t) = h_0(t) \exp & (1,102(Usia) - 0,512(Status\ Anemia\ 2)) \\ & + 1,469(Tekanan\ Darah\ Sistolik\ 2)) + 1,994(Tekanan\ Darah\ Sistolik\ 3)) \\ & + 0,545(Tekanan\ Darah\ Diastolik\ 2)) + 0,695(Kadar\ Ureum\ 2)) \\ & + 2,031(Kadar\ Ureum\ 3)) + 2,907(Kadar\ Ureum\ 4)) \\ & + 0,534(Kadar\ Kreatinin\ 2)) + 1,168(Kadar\ Kreatinin\ 3)) \\ & + 1,855(Kadar\ Kreatinin\ 4)) - 0,223(Kadar\ Hemoglobin\ 2)) \\ & - 0,084(Kadar\ Hemoglobin\ 3)) \end{aligned}$$

Sebelum dilakukan pemodelan dengan model MARS untuk analisis *survival*, maka perlu diketahui pola hubungan antara variabel prediktor dan variabel respon pada pemodelan *Cox* PH dalam model MARS.



Berdasarkan gambar diatas terlihat bahwa pola hubungan variabel respon (*martingale residual*) dan variabel prediktor (usia) menunjukkan pola hubungan yang tidak mengikuti suatu pola tertentu. Pola hubungan yang tidak cenderung membentuk pola tertentu, menunjukkan bahwa penelitian ini dapat digunakan dengan MARS.

Pemilihan model terbaik MARS yaitu dengan kriteria model yang memiliki GCV terkecil yaitu 0,341. Selanjutnya, penentuan model terbaik dilanjutkan dengan melihat nilai MSE terkecil. Nilai MSE terkecil adalah 0,295. Model MARS terbaik diperoleh dari model ke- dengan kombinasi BF sebesar 28, MI sebesar 3, dan MO sebesar 0. Berdasarkan kombinasi tersebut maka didapatkan model MARS terbaik sebagai berikut:

$$\hat{Y} = -0,144 + 0,159 * BF_1 + 0,053 * BF_{21} + BF_{21} + 0,051 * BF_{27}$$

Dimana :

$$BF_1 = \max(0, X_1 - 64);$$

$$BF_2 = \max(0, 64 - X_1);$$

$$BF_{20} = \max(X_7 = 1 \text{ OR } X_7 = 3) * BF_2;$$

$$BF_{21} = \max(X_3 = 2) * BF_{20};$$

$$BF_{24} = \max(0, 58 - X_1);$$

$$BF_{26} = \max(X_3 = 1 \text{ OR } X_3 = 3) * BF_{24};$$

$$BF_{27} = \max(X_7 = 1 \text{ OR } X_7 = 3) * BF_{26};$$

Sehingga pemodelan *cox* PH dengan model MARS adalah sebagai berikut:

$$\hat{h}(t) = h_0(t) \exp(0,144 + 0,159 * BF_1 + 0,053 * BF_{21} + BF_{21} + 0,051 * BF_{27})$$

Model MARS terbaik dengan variabel respon martingale residual dapat diketahui bahwa variabel yang masuk dalam model adalah usia (X_1), tekanan darah sistolik (X_3), kadar ureum(X_5), dan kadar hemoglobin(X_7).

KESIMPULAN

Berdasarkan hasil pemodelan *Cox Proportional Hazard* dengan pendekatan MARS, yang digunakan adalah kombinasi Basis Fungsi, Maksimum Interaksi dan Minimum Observasinya adalah 28, 3, dan 0 dengan nilai GCV minimum adalah 0,341 dan nilai MSE terkecil yaitu 0,295. Variabel yang masuk dalam model adalah usia (X_1), tekanan darah sistolik (X_3), kadar ureum(X_5), dan kadar hemoglobin(X_7).

DAFTAR PUSTAKA

- Brunner dan Suddarth. 2008. *Buku Ajar Keperawatan Medikal Bedah*. Jakarta : EGC.
- Collet, D. 2003. *Modeling Data in Medical Research*. London : Chapman Hall/CRC.
- Eubank, R.L. 1998. *Spline Smoothing and Nonparametric Regression*. New York : Marcel Dekker.
- Friedman, J.H. 1991. *Multivariate Adaptive Regression Splines*. California : Stanford University Press.
- Indonesia Renal Registry 2017*.
- Hosmer, D., Lemeshow, S., dan May, S. 2008. *Applied Survival Analysis*. New Jersey : John Wiley Sons.
- Kleinbaum, D.G., dan Klein, M. 2005. *Survival Analysis : A self-learning*. USA : Springer.
- Kriner, M. 2007. *Survival Analysis with Multivariate Adaptive Regression Splines*. Disertasi. Muenchen University.
- Nisa dan Budiantara. 2012. *Analisis Survival dengan Pendekatan Multivariate Adaptive Regression Splines pada Kasus Demam Berdarah Dengue (DBD)*. Jurnal Sains dan Seni ITS. Vol. 1, No. 1, 318-323
- Price dan Wilson. 2006. *Patofisiologi Konsep Klinis Proses-proses Penyakit*. Jakarta : EGC.

